Iranian Journal of Animal Science Research
Homepage: http://ijasr.um.ac.ir

Research Article
Vol. 16, No.2, 2024, p. 249-258

Genome-Wide Association Study Based on Mega-Analysis to Detect Genomic Regions Associated with Prolificacy in Sheep

Hossein Mohammadi[1*]

1- Assistant Professor, Department of Animal Science, Faculty of Agriculture, Arak University, Arak, Iran
*Corresponding Author's Email: H-mohammadi37@araku.ac.ir

Received: 24-07-2023
Revised: 30-09-2023
Accepted: 21-10-2023
Available Online: 21-10-2023

How to cite this article:
Mohammadi, H. (2024). Genome-wide association study based on mega-analysis to detect genomic regions associated with Prolificacy in sheep. *Iranian Journal of Animal Science Research*, *16*(2), 249-258 . (In Persian with English abstract). http://doi.org/10.22067/ijasr.2023.83596.1164

Introduction: Genetic architecture of sheep reproduction is increasingly gaining scientific interest due to the major impact on sheep production systems. The number of lambs per lambing is one of the m ost important reproductive traits in sheep. Many studies have reported that genetic mechanisms play an important role in the variation of litter size in sheep. Reproductive traits normally show low heritability and therefore response to conventional selection methods is not satisfactory for these traits. Considering the genetic information of the genetic variants underlying reproduction variability could efficiently increase the selection efficacy. Genome-wide association studies (GWAS) have been used to identify associations between genotypes and phenotypes as well as candidate genes for reproductive economically important traits. Statistical power in GWAS is mostly affected by sample size. The low sample size is hence a main obstacle in GWAS. Combining multiple data sets of different studies for joint (mega) GWAS provides an opportunity to increase the sample size required for GWAS. This study was performed to identify genomic regions affecting litter size in different sheep breeds using the mega-analysis of GWAS.

Materials and Methods: Multi-population joint GWAS was performed using genotypic and phenotypic data of three sheep breeds including native Zandi and two breed retrieved from the database. Quality control was performed using the Plink software. The markers or individuals were removed from the further study based on the following criteria: (1) unknown chromosomal or physical location, call rate <0.90, missing genotype frequency >0.05, minor allele frequency (MAF) < 0.05, and a P value for Hardy–Weinberg equilibrium test less than 10^{-6}. Before analysis, imputation of missing genotypes for combined data set was implemented by LD-kNNi method. Mega-analysis was performed using a mixed linear model in TASSEL software considering kinship and population structure (top five components of principal component analysis (PCA)) as confounding effects. The quantile–quantile (Q–Q) plot was visualized by plotting the distribution of obtained vs. expected log10 (P-value). The association results along the genome and the significant SNPs were visualized in the Manhattan plot. To account for multiple test problem and identify the genome-wide significance level, Bonferroni test was used based on the number of independent SNPs obtained from pairwise linkage disequilibrium analysis. After GWAS analysis, the 500 bp sequence upstream and downstream of the significant SNP was explored to identify the adjacent candidate genes using *ARS-UI_Ramb_v2.0* (Genome Data Viewer).

Results and Discussion: In the present study, we implemented a mega GWAS using three different sheep breed data to identify the genetic mechanisms responsible for litter size in sheep. After quality control, 671 animals and 45167 SNP markers were kept for further analysis. The results of the mega-analysis identified nine marker on chromosome on chromosomes 1 (two SNP), 2, 3 (two SNP), 10, 13 (two SNP), and 22. The quantile–quantile plot that features the total distribution of the observed P-values (−log10 P-values) of quality passed SNPs vs. the expected values, showed the effective control for confounding effects. Many of the significant SNPs identified in this study were located in or very adjacent to known genes (*DLG1*, *CLSTN2*, *INHBE*, *TCFL5*, and *RBP4*) that have been already reported for their contribution to fertility and pregnancy success. It has been

http://doi.org/10.22067/ijasr.2023.83596.1164

reported that the *RBP4* gene is expressed during the period of fast elongation of the pig blastocyst which is a crucial period for the survival of the embryos. Also, it has been suggested that *CLSTN2* has the main contribution in uterine and conceptus physiology during the establishment of pregnancy and therefore can be considered as a candidate gene for litter size. *INHBE* has an essential function during ovulation and pregnancy through extracellular matrix (ECM) components degradation and therefore enabling cell migration and angiogenesis.

Conclusion: Comparison of the results of this study with previous reports showed that the mega-analysis of GWAS, compared to the meta-analysis already reported for GWAS results, had comparable power in identifying genomic regions influencing litter size in sheep but identified fewer genomic regions than individual GWAS for each breed. No previously reported major genes controlling litter size in sheep were identified using our mega GWAS. The results of our research are suggested for further investigations in identifying causal genetic variants or genomic regions underlying the litter size variation in sheep and can be used to understand the genetic mechanism controlling this trait.

Keywords: Association analysis, Genomic region, Genetic Marker, Improving fertility, Mega-Analysis.

نشریه پژوهشهای علوم دامی ایران
Homepage: http://ijasr.um.ac.ir

مقاله پژوهشی
جلد ۱۶، شماره ۲، تابستان ۱۴۰۳، ص. ۲۵۸-۲۴۹

پویش ژنومی بر پایه آنالیز چند جمعیتی مشترک مرتبط با صفت تعداد نتاج متولد شده در گوسفندان نژاد بومی و خارجی

حسین محمدی[1*]

تاریخ دریافت: ۱۴۰۲/۰۵/۰۲
تاریخ بازنگری: ۱۴۰۲/۰۷/۰۸
تاریخ پذیرش: ۱۴۰۲/۰۷/۲۹

چکیده

تعداد بره متولد شده در هر زایش یکی از مهم‌ترین صفات اقتصادی و تولیدمثلی در گوسفند است. هدف از پژوهش حاضر، شناسایی مناطق ژنومی و ژن‌های کاندیدای مرتبط با چندقلوزایی در نژادهای مختلف گوسفند با رویکرد مگاآنالیز پویش ژنومی از طریق استفاده از اطلاعات مربوط به سه نژاد زندی، راهمنی و کایاس می‌باشد. بدین منظور، از اطلاعات ژنوتیپی و فنوتیپی ۶۸۲ رأس دام شامل نژادهای زندی (۹۶)، راهمنی (۴۸) و کایاس (۵۳۸) تعیین ژنوتیپ شده با از آرایه‌های 50K گوسفندی، استفاده شد. پس از مراحل مختلف کنترل کیفیت و ادغام داده‌های تعیین ژنوتیپ شده، ۶۷۱ رأس دام و ۴۵۱۶۷ نشانگر SNP برای ادامه آنالیزهای پویش ژنومی مورد استفاده قرار گرفتند. مگاآنالیز با استفاده از مدل خطی مختلط در نرم‌افزار TASSEL با در نظر گرفتن روابط خویشاوندی و ساختار جمعیتی انجام شد. نتایج حاصل نشان داد که تعداد نه نشانگر روی کروموزوم‌های شماره ۱ (دو نشانگر)، ۲، ۳ (دو نشانگر)، ۱۰، ۱۳ (دو نشانگر) و ۲۲ به‌طور معنی‌داری با صفت چندقلوزایی مرتبط می‌باشند. بررسی مناطق ژنومی کاندیدا به‌وسیله پایگاه‌های داده‌ای برخط نشان داد ژن‌های کاندیدای *DLG1*، *CLSTN2*، *INHBE*، *TCFL5* و *RBP4* نقش مؤثری در باروری، آبستنی موفق، فرآیند آزادسازی تخمک و اوولاسیون دارند. نتایج این پژوهش می‌تواند در درک ساز و کار ژنتیکی کنترل‌کننده چندقلوزایی در گوسفند مورد استفاده قرار گیرد.

واژه‌های کلیدی: آنالیز پیوستگی، بهبود باروری، مگاآنالیز، ناحیه ژنومی، نشانگر ژنتیکی

مقدمه

صفات تولیدمثلی، مهم‌ترین صفات مؤثر بر سودآوری اقتصادی سیستم‌های پرورش گوسفند هستند و سالانه به‌طور قابل ملاحظه‌ای تحت تأثیر توانایی میش در تولید بره‌های بیشتر قرار می‌گیرند (Abdoli *et al.*, 2019). از سوی دیگر، تولیدمثل یک فرآیند پیچیده زیست‌شناختی بوده که علاوه‌بر شرایط محیطی تحت تأثیر فاکتورهای ژنتیکی نیز قرار می‌گیرد. از طرفی، تعداد نتاج متولد شده در هر زایش یکی از مهم‌ترین صفت اقتصادی در صنعت پرورش گوسفند است. در واقع، یکی از اجزای بیولوژیکی در تولید، صفات تولیدمثلی می‌باشد، لذا بازده پرورش گوسفند به‌مقدار زیادی تابع توان تولید مثلی میش‌ها می‌باشد. به‌عبارت دیگر، افزایش تعداد بره‌های متولد شده به‌ازای هر میش در یک سال سبب افزایش بازده تولید در پرورش گوسفند خواهد شد (Zhang *et al.*, 2022).

گوسفند زندی یکی از نژادهای متوسط وزن کشور بوده و از نظر صفات عملکردی دارای خصوصیات مطلوبی می‌باشد، به‌طوری‌که چربی زیرپوستی در این نژاد در مقایسه با بسیاری از نژادهای گوسفند دیگر ایرانی کمتر می‌باشد. همچنین، مقاومت در برابر سرما و گرما (از ۲۰- درجه سانتی‌گراد در زمستان تا ۴۰ درجه سانتی‌گراد در تابستان)، شرایط سخت تغذیه‌ای و مراتع ضعیف و هوای بارانی و برفی در این گوسفند بسیار شایان توجه است (Bohlouli *et al.*, 2013). گوسفند نژاد راهمنی جزو سه نژاد اصلی پرورشی در کشور مصر است و نژاد مقاوم به شرایط آب و هوایی گرم و خشک بوده و به نژاد با تولید گوشت کیفیت بالا شهرت دارد (El-Halawany *et al.*, 2016). همچنین، نژاد کایاس مورد استفاده در پژوهش حاضر، یکی از نژادهای گوسفند معروف کشور یونان از نظر صفات تولیدی و تولیدمثلی می‌باشد. به‌طوری‌که این نژاد دارای تولید شیر بالا و عملکرد تولیدمثلی مطلوبی در سیستم‌های پرورشی صنعتی و نیمه صنعتی است (Tsartsianidou *et al.*, 2023).

۱- استادیار، گروه علوم دامی، دانشکده کشاورزی، دانشگاه اراک، اراک، ایران.
(*- نویسنده مسئول: (Email: H-mohammadi37@araku.ac.ir
http://doi.org/10.22067/ijasr.2023.83596.1164

در طی سال‌های اخیر، تکنیک‌های بیولوژی مولکولی با سرعت شگرفی توسعه پیدا کرده و موجب گردیده است تا پروژه‌های مختلف توالی‌یابی کل ژنوم، از جمله گوسفند تکمیل شود و متعاقب آن پانل-های مختلفی از آرایه‌های نانویی با تراکم متفاوت طراحی شده و انجام مطالعات پویش کل ژنومی را امکان‌پذیر نماید. مطالعه پویش ژنومی می‌تواند به شناسایی دقیق ژن‌های مرتبط با صفات اقتصادی کمک کرده و یافته‌های آن برای انتخاب به‌کمک نشانگر مفید باشد (Johnston *et al.*, 2011).

مطالعات مختلفی برای شناسایی مناطق ژنومی مؤثر بر صفات تولیدمثلی در گوسفند با رویکرد پویش کل ژنومی انفرادی انجام شده است (Xu *et al.*, 2018; Zhang *et al.*, Abdoli *et al.*, 2019; 2022; Ramos *et al.*, 2023). اندازه جمعیت مورد مطالعه یکی از پارامترهای اصلی مؤثر بر صحت پیش‌بینی ژنومی است. در مورد نژادهای با اندازه جمعیت کوچک، دستیابی به یک جمعیت پایه به اندازه کافی بزرگ، ممکن است بسیار دشوار و یا غیر ممکن باشد که در این صورت، اضافه کردن افراد از نژادهای دیگر می‌تواند به حل مشکل کمک نماید (Marjanovic and Calus, 2020). در آنالیز پویش ژنومی، اندازه نمونه مهم‌ترین عاملی است که قدرت آماری را تحت تأثیر قرار می‌دهد و در کنترل محقق است. به‌طور خاص، اندازه نمونه یک مانع بزرگ برای آنالیزهای پویش ژنومی در صفاتی است که اندازه‌گیری آن‌ها دشوار و یا هزینه‌بر است. یکی از راهکارهای حل محدودیت اندازه نمونه در فهم ساز و کار ژنتیکی صفات، ترکیب کردن نمونه‌های کوچک‌تر برای پویش ژنومی مشترک (مگاآنالیز) است (Gebreyesus *et al.*, 2019). ترکیب کردن مجموعه داده‌ها برای انجام آنالیز پویش ژنومی مشترک، یک روش کارآمد برای افزایش قدرت پویش ژنومی برای مطالعات در این زمینه است (Bouwman *et al.*, 2018). با در نظر گرفتن محدودیت‌های کمتر در ادغام نمونه‌های حیوانی، انتخاب روش‌های اجرای پویش ژنومی بر پایه رویکرد مگاآنالیز برای صفات تولیدمثلی می‌تواند خیلی پراهمیت باشد (Gebreyesus *et al.*, 2019).

اخیراً، پژوهشی با عنوان مطالعه پویش ژنومی بر پایه مگاآنالیز برای شناسایی جایگاه‌های ژنی مؤثر بر نرخ چندقلوزایی در شش نژاد مختلف گوسفند به تعداد ۴۶۲ رأس میش شامل نژادهای وادی، هوو، آیسلندیک، فینشیپ، رومانف و تکسل با هدف درک ساز و کار ژنتیکی این صفت با استفاده از مدل خطی مختلط انجام شده است که موفق به شناسایی ژن‌های کاندیدا شامل *OPCML*، *GULP1*، *RBP4*، *MMP2* و *LPCAT2* روی کروموزوم‌های شماره ۱، ۳، ۱۴، ۱۷ و ۲۲ شدند. نتایج نشان داد که روش پویش ژنومی بر مبنای مسیر کارآیی بالاتری برای یافتن مناطق ژنومی و درک بهتری از معماری ژنتیکی صفت تعداد نتاج متولد شده نسبت به آنالیز پویش ژنومی بر پایه تک نشانگری دارد (Gholizadeh and Esmaeili-Fard, 2022). مطالعه مروری دیگری در سال ۲۰۲۱ با هدف شناسایی ژن-های کاندیدای و مسیرهای زیستی مرتبط با تعداد بره متولد شده در گوسفند بر اساس آنالیز مسیر انجام شده است، که نتایج نشان داد، از میان ۲۱ ژن کاندیدای شناسایی شده ژن‌های *FLT1* و *CCL2* و از میان ۲۰ مسیر زیستی مرتبط با تعداد نتاج، مسیر زیستی negative regulation of vascular endothelial cell proliferation بیشترین ارتباط را با تعداد نتاج متولد شده در هر زایش را داشتند (Ghiasi and Abdollahi-Arpanahi, 2021).

با توجه به اینکه تا به امروز، اثر ترکیب کردن مجموعه داده‌های نژادهای مورد مطالعه برای انجام آنالیز پویش ژنومی مشترک در شناسایی مناطق ژنومی مؤثر بر چندقلوزایی گزارش نشده است. بنابراین، هدف از پژوهش حاضر انجام مگاآنالیز پویش ژنومی و شناسایی مناطق ژنومی مؤثر بر چندقلوزایی گوسفند با استفاده از ترکیب نمونه‌های سه نژاد شامل ۶۸۲ رأس دام بومی زندی (۹۶)، راهمنی مصر (۴۸) و کایاس یونان (۵۳۸) بود. همچنین عملکرد مگاآنالیز با گزارش‌های قبلی پویش ژنومی انفرادی روی این نژادها مورد مقایسه قرار گرفت.

مواد و روش‌ها

در پژوهش حاضر از اطلاعات ژنوتیپی و فنوتیپی سه نژاد مختلف گوسفند شامل نژادهای زندی (بومی)، راهمنی (مصری) و کایاس (یونانی) استفاده شد. ابتدا از ۲۰۰ گوسفند نژاد زندی حاضر در دو گله مرکز اصلاح نژاد خجیر واقع در تهران وابسته به سازمان جهاد کشاورزی استان تهران، مقدار ۷-۵ سی‌سی خون استحصال و همراه با ۰/۵ میلی‌لیتر EDTA در لوله‌های خلأ با pH ۸-۷/۵ نگهداری شد. از میان نمونه‌های اخذ شده پس از آنالیز شجره، حیواناتی که کمترین رابطه خویشاوندی را با هم داشته باشند، به تعداد ۹۶ رأس انتخاب شدند. تعدادی از میش‌ها (۴۴ رأس) دارای حداقل یک رکورد دوقلوزایی و تعدادی از میش‌ها (۵۲ رأس) میش دارای رکورد تک قلوزایی انتخاب شدند. استخراج DNA با استفاده از روش بهینه یافته استخراج نمکی از خون کامل انجام شد (Helms, 1990). ژنوتیپ نمونه‌ها در کمپانی Neogen (http://genomics.neogen.com) در کشور آمریکا با استفاده از آرایه‌های Illumina OvineSNP50K BeadChip تعیین ژنوتیپ شدند.

علاوه‌براین، در پژوهش حاضر از اطلاعات ژنوتیپی مجموع ۴۸ رأس گوسفندان غیر خویشاوند مصری مربوط به نژاد راهمنی استفاده گردید (El-Halawany *et al.*, 2016). لازم به ذکر است که تعیین ژنوتیپ نمونه‌ها در مرکز بین‌المللی تحقیقات کشاورزی مناطق خشک کشور مصر انجام شده بود. ژنوتیپ نمونه‌ها در کمپانی Neogen GeneSeek در کشور آمریکا با استفاده از آرایه‌های Illumina OvineSNP50K BeadChip با به‌کارگیری پروتکل استاندارد

شرکت ایلومینا انجام شده بودند. داده‌های مورد استفاده در این پژوهش در پایگاه برخط Dryad (https://datadryad.org) با شماره دسترسی 10.5061/dryad.k83b3 در دسترس می‌باشد. همچنین، از اطلاعات ژنوتیپی و فنوتیپی مجموع ۵۳۸ رأس گوسفندان غیر خویشاوند یونانی مربوط به نژاد کایاس[1] نیز استفاده گردید (Tsartsianidou *et al.*, 2023). نمونه‌ها با استفاده از آرایه‌های شرکت ایلومینا Ovine SNP50 Bead Chip برای ۵۴۲۴۱ جایگاه نشانگری تعیین ژنوتیپ شده بودند. داده‌های مورد استفاده در این پژوهش در پایگاه برخط Mendeley (https://data.mendeley.com/datasets) با شماره دسترسی nvjn8mvwtg/1 در دسترس می‌باشد.

برای اطمینان از کیفیت داده‌های حاصل از تعیین ژنوتیپ در آنالیزهای نهایی مراحل مختلف کنترل کیفیت بر روی داده‌های اولیه اعمال شد. بدین منظور، در ابتدا نمونه‌هایی که فراوانی ژنوتیپ در آن‌ها کمتر از ۹۰ درصد بود، شناسایی و حذف شدند. در مرحله بعد، نشانگرهایی که فراوانی آلل نادر (MAF[2]) در آن‌ها کمتر از پنج درصد بود، حذف شدند. سپس، نشانگرهای SNPای که فراوانی ژنوتیپ آن‌ها در نمونه‌ها کمتر از ۹۵ درصد بود، شناسایی و حذف شدند. همچنین برای SNPهای باقی‌مانده آن‌هایی که در تعادل هاردی-واینبرگ قرار نداشتند، با سطح احتمال 10^{-6} شناسایی و حذف شدند. مراحل مختلف غربالگری با استفاده از نرم‌افزار PLINK (نسخه ۱/۹۰) انجام شد (Purcell *et al.*, 2007).

پس از کنترل کیفیت، داده‌های پیرایش شده با استفاده از نرم‌افزار PLINK ادغام شدند و دور بعدی کنترل کیفیت روی داده‌های ادغام شده با معیارهای مشابه دور قبل انجام شد. در مرحله بعد، قبل از انجام آنالیز پویش ژنومی، ایمپیوتیشن با روش LD-kNNi (Money *et al.*, 2015) انجام شد. ادغام داده‌های جمعیت‌های انفرادی در قالب یک مجموعه داده بزرگ، امکان مگاایمپیوتیشن و مگاآنالیز (Fritsche *et al.*, 2016) و شناسایی واریانت‌های بیشتر نسبت به متاایمپیوتیشن را فراهم می‌کند (Gorski *et al.*, 2019). ایمپیوتیشن با استنتاج SNPهای گم‌شده یا تعیین ژنوتیپ نشده، ابزاری مفید برای آنالیز پویش ژنومی است که می‌تواند مجموعه SNPهای مشابه برای افراد مورد مطالعه را فراهم کند و بنابراین، منجر به افزایش قدرت آزمون پویش ژنومی می‌گردد (Gorski *et al.*, 2019).

در مرحله بعد برای بررسی ساختار جمعیتی و مدل کردن آنالیز پویش ژنومی از مدل خطی مختلط در نرم‌افزار TASSEL 5 (Bradbury *et al.*, 2007) استفاده گردید. دو عامل مختل‌کننده که می‌توانند نتایج آنالیز پویش ژنومی را دچار اریبی کنند، شامل خویشاوندی و ساختار لایه‌بندی جمعیتی هستند. مدل‌های خطی عمومی امکان در نظر گرفتن لایه‌بندی جمعیتی را از طریق وارد کردن ابعاد مؤلفه‌های اصلی در مدل را فراهم می‌کنند. ولی نمی‌توانند خویشاوندی را در نظر بگیرند. مدل مختلط این امکان را فراهم می‌سازد که علاوه‌بر لایه‌بندی جمعیتی، خویشاوندی نیز در قالب ماتریس روابط خویشاوندی را در مدل وارد شود. برای بررسی ساختار جمعیت ادغام شده از تجزیه مؤلفه‌های اصلی (PCA[3]) استفاده شد. این تجزیه با استفاده از نرم‌افزار PLINK انجام شد و تعداد هشت مؤلفه اول ماتریس ژنوتیپی محاسبه شد.

مدل مورد استفاده بر پایه مدل خطی مختلط به‌شکل زیر بود:

$$y = W\alpha + x\beta + Zu + e \quad \text{(۱) معادله}$$

که در آن، **y**: بردار مشاهدات فنوتیپی، **W**: ماتریس ضرایب مربوط به اثر ثابت شامل نژاد، سن و پنج سطح اول تجزیه مؤلفه اصلی با بیشترین واریانس توجیه شده، ***x***: بردار ژنوتیپ‌های نشانگر تحت آزمون، **u**: بردار اثرات تصادفی پلی‌ژنیک، **Z**: ماتریس روابط ژنومیک، **e**: بردار اثر تصادفی باقی‌مانده‌های، **α** و **β**: بردار ضرایب ارتباط دهنده است.

برای تصحیح آزمون‌های چندگانه و رسیدن به سطح معنی‌داری ژنومی از آزمون بنفرونی بر اساس تعداد SNPهای مستقل استفاده شد. برای این منظور، با استفاده از نرم‌افزار PLINK و دستور indep-pairwise 50 5 0.05 استفاده گردید. برای در نظر گرفتن سطح معنی‌داری پنج درصد ژنومی (α) و با تعداد SNPهای مستقل (N)، آستانه معنی‌داری ژنومی برآورد شد (0.05/11812). همچنین، برای بررسی برازش مدل و نتایج از پلات‌های Q-Q و منهتن استفاده شد و برای ترسیم گراف‌های مربوطه از نرم‌افزار SNPEVG2 استفاده گردید (Wang *et al.*, 2012).

بررسی ژن‌های گزارش شده در مناطق کاندیدا و ۵۰۰ کیلو جفت‌باز اطراف آن‌ها از پایگاه اطلاعاتی برخط NCBI به‌وسیله Genome Data Viewer (https://www.ncbi.nlm.nih.gov/genome/gdv?org=ovis-aries&group=bovidae) بر اساس آخرین نسخه اسبلی گوسفند (*ARS-UI_Ramb_v2.0*) مورد بررسی قرار گرفت. جهت تعیین طبقات عملکردی ژنی و مسیرهای متابولیکی ژن‌های شناسایی شده از پایگاه اطلاعاتی PANTHER (http://www.pantherdb.org) استفاده گردید. همچنین، برای تفسیر بهتر عملکرد ژن‌های به‌دست آمده از پایگاه‌های اطلاعاتی آنلاین GeneCards (http://www.genecards.org) و UniProtKB (http://www.uniprot.org) استفاده شد.

1- Chios
2- Minor allele frequency
3- Principal component analysis

جدول ۱- توصیف مراحل کنترل کیفیت داده‌های مربوط به نژادهای مختلف مورد بررسی در سطح افراد

Table 1- Description of the quality control steps in different studied breeds at the level of individuals

	زندی Zandi	کایاس Chios	راهمنی Rahmani
تعداد حیوانات Number of animals	96	538	48
حذف حیواناتی با نرخ فراخوانی کمتر از ۹۰ درصد Excluding animals with call rate<0.90	0	3	2
حذف حیواناتی با یکسانی بر اساس موقعیت افراد بالاتر از ۹۹ درصد Excluding animals with IBS>0.99	0	3	1
حذف حیواناتی که در آنالیز PCA خارج از توده جمعیتی خود قرار داشتند Excluding animals that were clustered outside of their population mass in PCA analysis	2	0	0
تعداد حیوانات باقی‌مانده Remaining animals	94	532	45

جدول ۲- خلاصه مراحل مختلف کنترل کیفیت نشانگرهای SNP برای آنالیزهای پویش ژنومی

Table 2- Summary of quality control steps for SNP markers used for GWAS analysis

مراحل مختلف کنترل کیفی Quality control filters	تعداد نشانگر SNP SNP number
تعداد نشانگرهای SNP قبل از کنترل کیفیت SNP number before QC	37844
حذف SNPهایی با فراوانی آللی نادر کمتر از ۵ درصد Excluding SNPs ≤5% MAF over all animals	1025
حذف SNPهایی با نرخ فراخوانی کمتر از ۹۵ درصد در کلیه حیوانات Excluding SNPs ≤95% Call rate over all animals	842
حذف SNPهای خارج از تعادل هاردی-واینبرگ با ارزش (۰/۰۰۰۰۰۱) Excluding SNPs with deviation from HW (<0.000001)	95
تعداد نشانگرهای SNP باقی‌مانده Remaining SNPs	35882
تعداد نشانگرهای SNP باقی‌مانده بعد از ایمپیوتیشن Remaining SNPs after imputation	45167

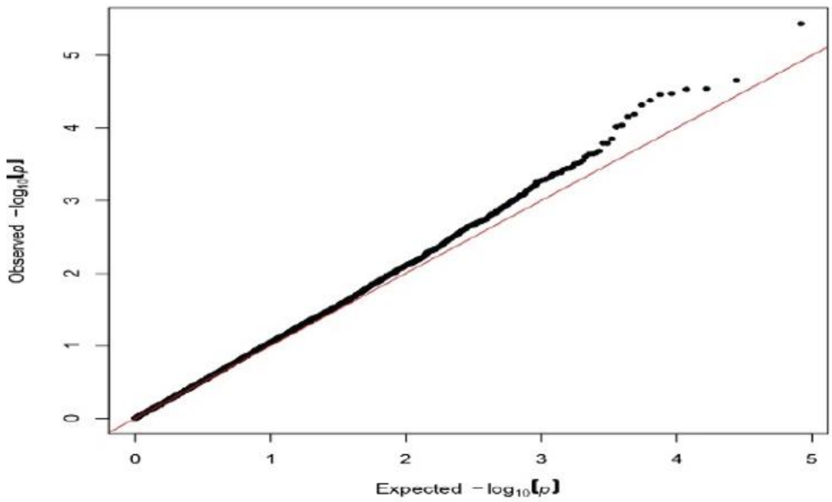
نتایج و بحث

نتایج مربوط به مراحل مختلف ویرایش داده‌های ژنوتیپ شده در جدول ۱ ارائه شده است. تعداد چهار حیوان با داشتن IBS[1] حدود ۰/۹۹ تا یک با احتمال تکراری بودن نمونه‌ها و تعیین ژنوتیپ چند باره یک حیوان و همچنین حیوانات خارج از توده نژادی خود در آنالیز PCA به‌منظور جلوگیری از ایجاد اریبی در محاسبات حذف شد (جدول ۱). پس از حذف داده‌های حیوانات حذف شده، داده‌های نژاد بومی زندی با نژادهای گوسفند مصری و یونانی با هم ترکیب شدند که دارای ۳۷۸۴۴ نشانگر SNP مشترک بودند که پس از مرحله کنترل کیفیت در سطح SNP در نهایت، تعداد ۳۵۸۸۲ نشانگر SNP باقی ماندند که شامل اطلاعات ۲۶ جفت کروموزوم گوسفند بودند، به‌طوری‌که پس از پایان ایمپیوتیشن تعداد ۶۷۱ دام و ۴۵۱۶۷ نشانگر SNP با متوسط MAF برابر با ۰/۳۰۴ در آنالیز پویش ژنومی مورد استفاده قرار گرفتند (جدول ۲).

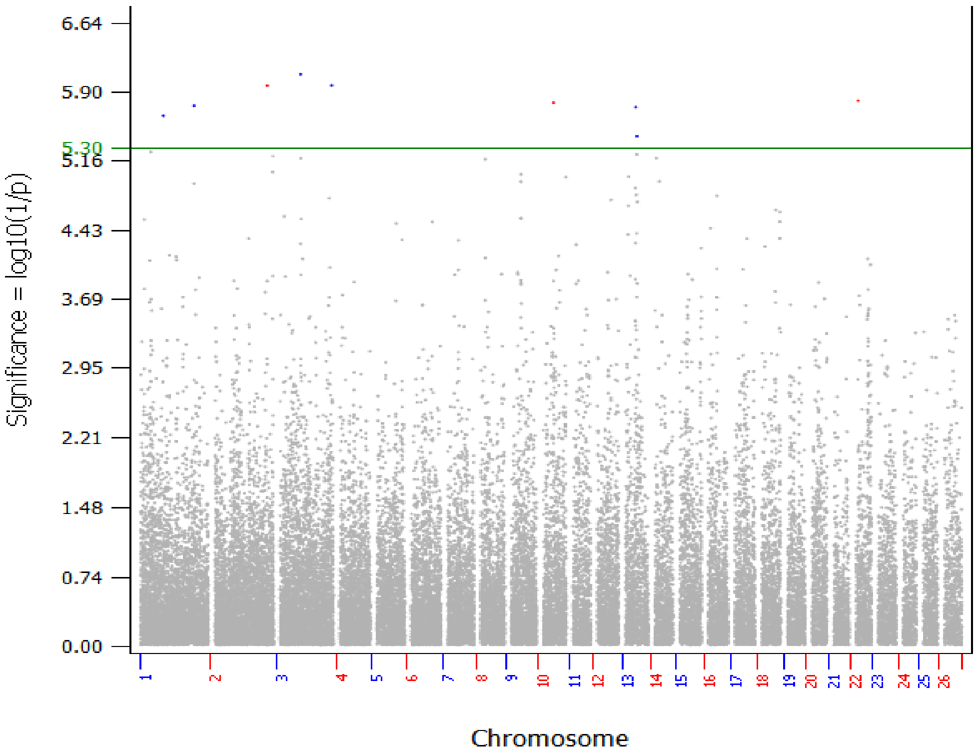
نتیجه تجزیه مؤلفه‌های اصلی نشان داد که سطح یک مؤلفه‌های اصلی در برابر سطح دو جمعیت‌ها را از هم تفکیک کرد. همچنین، پنج سطح اول که بالاترین میزان واریانس ژنتیکی را توجیه می‌کردند، برای کنترل ساختار جمعیتی در مدل آماری گنجانده شدند. پلات Q-Q و منهتن مرتبط با صفت تعداد نتاج متولد شده در هر زایش در شکل ۱ و ۲ ارائه شده است. چندین روش برای تخمین کنترل تورم ژنومیکی (λ) وجود دارد که در این تحقیق روش تخمین‌گر میانه اجرا شد. فاکتور تورم کنترل جمعیتی بزرگ‌تر از یک (λ>1)، وجود لایه-بندی جمعیتی و یا خطای تعیین ژنوتیپ را نشان می‌دهد. فاکتور لامبدا از طریق تجزیه و تحلیل پیوستگی در PLINK برای صفات مورد مطالعه محاسبه شد که برای صفت مورد مطالعه، ۱/۰۲۴ بود که همان‌طور که مشاهده می‌شود، تقریباً برابر با یک بودند و نشان‌دهنده

1- Identical by state

عدم وجود لایه‌بندی جمعیتی برای اجرای تجزیه و تحلیل‌های پویش کل ژنومی بود. در مطالعاتی که لامبدا کوچک‌تر از ۱/۱ یا مساوی یک است، نیازی به تصحیح اثرات ساختار جمعیتی نیست.



شکل ۱– پلات Q-Q نتایج مگاآنالیز پویش ژنومی برای کنترل کیفیت نتایج
Figure 1- Q-Q plot of mega GWAS results for assessing quality of results



شکل ۲- پلات منهتن نتایج مگاآنالیز پویش ژنومی با استفاده از مدل خطی مختلط برای شناسایی نواحی ژنومی مرتبط با چندقلوزایی در گوسفند در سطح معنی‌داری ژنومی
Figure 2- Manhattan plot of mega GWAS results obtained from linear mixed model to identify genomic regions associated with prolificacy in sheep with genome-wide significance level

نتایج مگاآنالیز پویش ژنومی، نه عدد نشانگر SNP در سطح ژنومی شناسایی کرد که به‌طور معنی‌داری در سطح ژنوم با میانگین چندقلوزایی در ارتباط بودند (شکل ۲). نشانگرهای معنی‌دار به‌ترتیب روی کروموزوم‌های شماره ۱، ۳ و ۱۳ هرکدام دو نشانگر و

کروموزوم‌های ۲، ۱۰ و ۲۲ هرکدام یک نشانگر قرار داشتند. بررسی‌های بیشتر در پایگاه‌های داده نشان داد که برخی از ژن‌های موجود در نواحی ژنومی شناسایی شده نقش مؤثری در عملکرد تولیدمثلی دارند که در این بخش توضیحاتی در ارتباط با نقش‌های فیزیولوژیکی این ژن‌ها ارائه خواهد شد.

نتایج تحقیق حاضر نشان داد که دو نشانگر در نواحی ژنومی ۱۸۹/۸ و ۲۴۶/۹ مگابازی روی کروموزوم شماره ۱ به‌طور معنی‌داری با چندقلوزایی ارتباط دارند. تجزیه بیشتر در پایگاه‌های داده نشان داد که این نشانگرها با ژن‌های *CLSTN2* و *DLG1* در ارتباط هستند. ژن *CLSTN2* در فرآیند تخمک‌اندازی و تبدیل سلول‌های تکا و گرانولوزا به سلول‌های لوتئال بزرگ و کوچک در جسم زرد می‌باشد (Hong *et al.*, 2006). در مطالعه پویش ژنومی در گوسفندان نژاد Pelibuey با هدف شناسایی ژن‌های کاندیدای مرتبط با دوقلوزایی ژن *CLSTN2* گزارش شده است (Hernández-(Montiel *et al.*, 2020. همچنین ژن *DLG1* نقش مؤثری در فعال‌سازی مسیر سیگنال‌دهی Hippo دارد که نقش اساسی در فرآیند آزادسازی تخمک از طریق افزایش سطح ۱۷ بتا استرادیول (E2) دارد (Ye *et al.*, 2017).

ژن *GULP1* روی کروموزوم شماره ۲ در ناحیه ژنومی ۱۱۹/۷ مگابازی قرار دارد. نشان داده شده است که این ژن با طول دوره لوتئال در گاو در ارتباط است (Tenghe *et al.*, 2016). همچنین این ژن تنظیم‌کننده کلیدی سیگنال‌دهی فاکتور رشد تغییردهنده بتا می‌باشد. ژن *GULP1* شناسایی شده در پژوهش حاضر در مطالعه قبلی آنالیز پویش ژنومی انفرادی مرتبط با نژاد کایاس گزارش شده است که در سطح ژنومی با چندقلوزایی در این نژاد مرتبط بود (Tsartsianidou *et al.*, 2023).

ژن *INHBE* روی کروموزوم شماره ۳ در ناحیه ژنومی ۱۶۱/۹ مگابازی قرار دارد. نقش‌های تنظیمی مهمی در لانه‌گزینی و تشکیل جفت برای باروری موفق ایفا می‌کند. نتایج مطالعات نشان داده است که چندشکلی در این ژن با صفات باروری در گاوهای پرورش یافته در مناطق گرمسیری در ارتباط است (Fortes *et al.*, 2023). این ژن که در فولیکول بالغ بیان می‌شود، دیواره فولیکولی پستانداران را طی اوولاسیون تخریب می‌کند و بنابراین، فعالیت این ژن به لحاظ ژنتیکی برای اوولاسیون مورد نیاز است (Lehnert and Reverter, 2013).

ژن *RBP4* روی کروموزوم شماره ۲۲ در ناحیه ژنومی ۱۴/۷ مگابازی به‌عنوان ژن کاندیدا گزارش شده است. این ژن طی فاز توسعه سریع بلاستوسیت بیان می‌شود که دوره حیاتی برای زنده‌مانی جنین است. این ژن، پروتئینی را کد می‌کند که در رحم و در اوایل آبستنی حضور دارد. این پروتئین‌ها به رتینول متصل می‌شوند، سپس رتینول متصل شده به‌وسیله سلول‌ها پذیرفته شده و جنین‌زایی آغاز می‌شود (Messer *et al.*, 1996).

جدول ۳– نتایج مگاآنالیز پویش ژنومی با مدل خطی مختلط برای شناسایی نواحی ژنومی و ژن‌های موجود در این مناطق ژنومی مرتبط با چندقلوزایی در گوسفند

Table 3– Mega GWAS results obtained from linear mixed model to identify genomic regions and genes associated with prolificacy in sheep

شماره کروموزوم Number of chromosomes	نشانگر ID SNP ID	آلل ۱ A1	آلل ۲ A2	حداقل فراوانی آلل نادر MAF	موقعیت ژنومی (جفت باز) Genomic position (bp)	ژن کاندیدای مرتبط با تولیدمثل Candidate gene related to reproductive
1	OAR1_204970872.1	T	C	0.361	189355910-190355910	*DLG1*
1	s09883.1	G	A	0.212	246413454-247413454	*CLSTN2*
2	OAR2_119795639.1	A	G	0.373	119289298-120289298	*GULP1*
3	OAR3_173173378.1	C	T	0.395	161428466-162428466	*INHBE*
10	OAR10_90315139.1	C	A	0.421	82164637-83164637	*TNFSF13B*
13	s16241.1	C	T	0.369	53751020-54751020	*TCFL5*
22	OAR22_18241256.1	G	A	0.283	14242914-15242914	*RBP4*

ترکیب کردن مجموعه‌های داده‌های چند جمعیت چالش‌های مخصوص به خود را دارد. یک از موانع اصلی عدم تجانس نمونه‌های جمعیت‌های مختلف است. این عدم تجانس‌ها ممکن است به علت‌های مختلف از جمله فاصله ژنتیکی بین جمعیت‌ها، تفاوت اندازه‌گیری صفات، محیط‌های پرورشی مختلف و آرایه‌های تعیین ژنوتیپ متفاوت باشد (Begum *et al.*, 2012). نتایج به‌دست آمده از مطالعات مختلف

آنالیز پویش ژنومی با استفاده از ادغام چندین مجموعه داده، نواحی ژنومی بیشتر و دقیق‌تری را در مقایسه با تجزیه انفرادی جمعیتی و همچنین فراتحلیل آن جمعیت‌ها گزارش کرده‌اند (Gebreyesus *et al.*, 2019).

تاکنون تعداد معدودی آنالیز پویش ژنومی با هدف شناسایی مناطق ژنومی مرتبط با صفت چندقلوزایی در نژادهای مختلف بومی در ایران انجام شده است که منجر به شناسایی و معرفی ژن‌های کاندیدای مختلفی شده است. در گوسفند نژاد بلوچی نشان داده شده است که واریانت‌های متفاوت ژن‌های *NTRK2*، *CTH*، *ANKRD13C*، *SRSF11*، *PTGER3*، *LDHB*، *LRRC40* و *KCNMA1* در تنوع موجود در صفت دوقلوزایی در این نژاد نقش دارند (Esmaeili-Fard *et al.*, 2021). همچنین، در مطالعه پویش ژنومی انجام شده روی گوسفندان نژاد لری- بختیاری، ژن *LHCGR* به‌عنوان ژن کاندیدای مؤثر بر دو قلوزایی معرفی شده است (Abdoli *et al.*, 2018). تفاوت در نتایج مطالعات مختلف پویش ژنومی می‌تواند به‌دلیل تفاوت‌های بین نژادی، روش‌های مورد استفاده در تجزیه داده‌ها و اندازه نمونه تعیین ژنوتیپ شده و دارای فنوتیپ باشد. این موضوع نشان می‌دهد که ساز و کارهای زیستی متفاوتی در تنوع موجود در صفت دوقلوزایی در بین نژادهای مختلف وجود دارد و می‌توان با آنالیز پویش ژنومی، ژن‌های کاندیدای مؤثر بر این صفت مهم اقتصادی را شناسایی کرد و از مگاآنالیز در ترکیب نتایج بهره گرفت.

نتیجه‌گیری کلی

در مجموع، نتایج حاصل از مگاآنالیز منجر به شناسایی نه نشانگر SNP در سطح ژنوم گردید که به‌طور معنی‌داری با چندقلوزایی در گوسفند در ارتباط بودند. بررسی عملکرد ژن‌های موجود در این مناطق ژنومی نشان داد که ژن‌های *DLG1*، *CLSTN2*، *INHBE*، *TCFL5* و *RBP4* نقش مؤثری در باروری و موفقیت آبستنی دارند و می‌توانند در برنامه‌های اصلاح نژادی مورد استفاده قرار گیرند.

References

1. Abdoli, R., Mirhoseini, S. Z., Ghavi Hossein-Zadeh, N., Zamani, P., & Gondro C. (2018). Genome-wide association study to identify genomic regions affecting prolificacy in Lori-Bakhtiari sheep. *Animal Genetics*, *49*(5), 488-491. https://doi.org/10.1111/age.12700.
2. Abdoli, R., Mirhoseini, S. Z., Ghavi Hossein-Zadeh, N., Zamani, P., Moradi, M. H., Ferdosi, M. H., Gondro, C., (2019). Genome-wide association study of first lambing age and lambing interval in sheep. *Small Ruminant Research*, *178*, 43–45. https://doi.org/10.1016/j.smallrumres.2019.07.014
3. Begum, F., Ghosh, D., Tseng, G. C., & Feingold, E. (2012). Comprehensive literature review and statistical considerations for GWAS meta-analysis. *Nucleic Acids Research*, *40*, 3777–3784. https://doi.org/10.1093/nar/gkr1255.
4. Bohlouli, M., Mohammadi, H., & Alijani, S. (2013). Genetic evaluation and genetic trend of growth traits of Zandi sheep in semi-arid Iran using random regression models. *Small Ruminant Research*, *114*, 195–201. https://doi.org/10.1016/j.smallrumres.2013.07.005
5. Bouwman, A. C., Daetwyler, H. D., Chamberlain, A. J., Ponce, C. Sargolzaei, M., Schenkel, F. S., Sahana, G., Govignon-Gion, A., Boitard, S., Dolezal, M., Pausch, H., Brøndum, R. F., Bowman, P. J., Thomsen, B., Guldbrandtsen, B., Lund, M. S., & Hayes, B. J. (2018). Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nature Genetics*, *50*, 362–367. https://doi.org/10.1038/s41588-018-0056-5.
6. Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, *23*(19), 2633-2635. https://doi.org/10.1093/bioinformatics/btm308.
7. El-Halawany, N., Zhou, X., Al-Tohamy, A. F., El-Sayd, Y. A., Shawky, A. E., Michal, J. J., & Jiang, Z. (2016). Genome-wide screening of candidate genes for improving fertility in Egyptian native Rahmani sheep. *Animal Genetics*, (4), 513. https://doi.org/10.1111/age.12437.
8. Esmaeili-Fard, S. M., Gholizadeh, M., Hafezian, S. H., & Abdollahi-Arpanahi, R. (2021). Genes and pathways affecting sheep productivity traits: Genetic parameters, genome-wide association mapping, and pathway enrichment analysis. *Frontiers in Genetics*, *1*, 1351. https://doi.org/10.3389/fgene.2021.710613.
9. Fortes, M. R. S., Reverter, A., Kelly, M., McCulloch, R., & Lehnert, S. A. (2013). Genome-wide association study for inhibin, luteinizing hormone, insulin-like growth factor 1, testicular size and semen traits in bovine species. *Andrology*, *1*, 644–650. https://doi.org/ 10.1111/j.2047-2927.2013.00101.x.
10. Fritsche, L. G., Igl, W., Bailey, J. N., Grassmann, F., Sengupta, S., Bragg-Gresham, J. L., & Heid, I. M. (2016). A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nature Genetics*, 48(2), 134-143. https://doi.org/10.1038/ng.3448.
11. Gebreyesus, G., Buitenhuis, A. J., Poulsen, N. A., Visker, M. H. P. W., Zhang, Q., van Valenberg, H. J. F., & Bovenhuis H. (2019). Combining multi-population datasets for joint genome-wide association and meta-analyses:

The case of bovine milk fat composition traits. *Journal of Dairy Science*, 102(12), 11124-11141. https://doi.org/10.3168/jds.2019-16676.
12. Gholizadeh, M., & Esmaeili-Fard, S. M. (2022). Multi-population joint genome-wide association study to detect genomic regions associated with litter size in sheep. *Animal Production Research, 11*(3), 15-26. (In Persian) https://doi.org/10.22124/AR.2022.21763.1688
13. Gorski, M., Günther, F., Winkler, T. W., Weber, B., & Heid, I. M. (2019). On the differences between mega- and meta-imputation and analysis exemplified on the genetics of age-related macular degeneration. *Genetic Epidemiology*, *43*(5), 559-576. https://doi.org/10.1002/gepi.22204.
14. Ghiasi, H., & Abdollahi-Arpanahi, R. (2021). The candidate genes and pathways affecting litter size in sheep. *Small Ruminant Research*, 205, 106546. https://doi.org/10.1016/j.smallrumres.2021.106546
15. Hernández-Montiel, W., Martínez-Núñez, M. A., Ramón-Ugalde, J. P., Román-Ponce, S. I., Calderón-Chagoya, R., & Zamora-Bustillos, R. (2020). Genome-wide association study reveals candidate genes for litter size traits in Pelibuey sheep. *Animals (Basel)*, *10*(3), 434. https://doi.org/10.3390/ani10030434.
16. Helms, C. (1990). Salting out procedure for human DNA extraction. retrieved april 20, 2010, from http://humgen.wustl.edu/hdk_lab_manual/dna/dna2.html
17. Hong, E. J., Park, S. H., Choi, K. C., Leung, P. C. K., & Jeung, E. B. (2006). Identification of estrogen-regulated genes by microarray analysis of the uterus of immature rats exposed to endocrine disrupting chemicals. *Reproduction Biological Endocrinology*, *4*, 1–12. https://doi.org/10.1186/1477-7827-4-49.
18. Johnston, S. E., McEwan, J. C., Pickering, N. K., Kijas, J. W., Beraldi, D., Pilkington, J. G., Pemberton, J. M., & Slate, J. (2011). Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular Ecology*, 20, 2555–2566. https://doi.org/10.1111/j.1365-294X.2011.05076.x.
19. Lehnert, S., & Reverter, T. (2013). Genome-wide association study of tropical composite bulls for reproduction traits. CSIRO Animal, Food and Health Sciences, Published by Meat & Livestock Australia Limited.
20. Marjanovic, J., & Calus, M. P. L. (2020). Factors affecting accuracy of estimated effective number of chromosome segments for numerically small breeds. *Journal of Animal Breeding and Genetics*, *138*, 151-160. https://doi.org/10.1111/jbg.12512.
21. Messer, L. A., Wang, L., Yelich, J., Pomp, D., Geisert, R. D., & Rothschild, M. F. (1996). Linkage mapping of the retinol binding protein 4 (RBP4) gene to porcine chromosome 14. *Mammalian Genome*, *7*, 396-410. https://doi.org/10.1007/s003359900117.
22. Money, D., Gardner, K., Migicovsky, Z., Schwaninger, H., Zhong, G. Y., & Myles, S. (2015). LinkImpute: fast and accurate genotype imputation for nonmodel organisms. *G3: Genes*, *Genomes*, *Genetics*, *5*(11), 2383-2390. https://doi.org/10.1534/g3.115.021667.
23. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., & Sham P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559-575. https://doi.org/10.1086/519795.
24. Ramos, Z., Garrick, D. J., Blair, H. T., Vera, B., Ciappesoni, G., & Kenyon, P. R. (2023). Genomic Regions associated with wool, growth and reproduction traits in Uruguayan Merino sheep. *Genes*, 14, 167. https://doi.org/10.3390/genes14010167.
25. Tsartsianidou, V., Pavlidis, A., Tosiou, E., Arsenos, G., Banos, G., & Triantafyllidis A. (2023). Novel genomic markers and genes related to reproduction in prolific Chios dairy sheep: A genome-wide association study. *Animal*, (3), 100723. https://doi.org/10.1016/j.animal.2023.100723.
26. Tenghe, A. M. M., Bouwman, A. C., Berglund, B., Strandberg, E., de Koning, D. J., & Veerkamp, R. F. (2016). Genome wide association study for endocrine fertility traits using single nucleotide polymorphism arrays and sequence variants in dairy cattle. *Journal of Dairy Science*, *99*(7), 5470-5485. https://doi.org/10.3168/jds.2015-10533.
27. Wang, S., Dvorkin, D., & Da, Y. (2012). SNPEVG: A graphical tool for GWAS graphing with mouse clicks. *BMC Bioinformatics*, *13*, 319. https://doi.org/10.1186/1471-2105-13-319.
28. Xu, S. S., Gao, L., Xie, X. L., Ren, Y. L., Shen, Z. Q., Wang, F., & Li, M. H. (2018). Genome-wide association analyses highlight the potential for different genetic mechanisms for litter size among sheep breeds. *Frontiers in Genetics*, *9*, 118. https://doi.org/10.3389/fgene.2018.00118.
29. Ye, H., Li, X., Zheng, T., Hu, C., Pan, Z., Huang, J., Li, J., Li, W., & Zheng, Y. (2017). The hippo signaling pathway regulates ovarian function via the proliferation of ovarian germline stem cells. *Cell Physiological Biochemistry*, *41*, 1051–1062. https://doi.org/10.1159/000464113
30. Zhang, Z., Sui, Z., Zhang, J., Li, Q., Zhang, Y., Wang, C., Li, X., & Xing, F. (2022). Identification of signatures of selection for litter size and pubertal initiation in two sheep populations. *Animals*, *12*, 2520. https://doi.org/10.3390/ani12192520.